Noise Reduction Using an Undecimated Discrete Wavelet Transform*

M. Lang (Member IEEE)
H. Guo (Student member IEEE)
J. E. Odegard (Student member IEEE)
C. S. Burrus (Fellow IEEE)
Department of Electrical and Computer Engineering
Rice University, Houston, TX 77251-1892

R. O. Wells, Jr.
Department of Mathematics
Rice University, Houston, TX 77251-1892

October 5, 1995

Paper to appear in IEEE Signal Processing Letters, 1995

Abstract

A new nonlinear noise reduction method is presented that uses the discrete wavelet transform. Similar to Donoho and Johnstone, we employ thresholding in the wavelet transform domain but, following a suggestion by Coifman, we use an undecimated, shift-invariant, nonorthogonal wavelet transform instead of the usual orthogonal one. This new approach can be interpreted as a repeated application of the original Donoho and Johnstone method for different shifts. The main feature of the new algorithm is a significantly improved noise reduction compared to the original wavelet based approach, both the l_2 error and visually, for a large class of signals. This is shown both theoretically as well as by experimental results.

1 Introduction

Recently a powerful approach for noise reduction has been proposed by Donoho and Johnstone [3, 4, 5]. It employs thresholding in the wavelet domain and has been shown to be asymptotically near optimal for a wide class of signals corrupted by additive white Gaussian noise. It has been successfully applied to 1D and 2D data such as NMR spectra, geophysical data, and to coherent imaging process data, e.g. synthetic aperture radar [7]. Moreover, the same method can be

*This research was supported by the Alexander von Humboldt foundation, ARPA, BNR, Texas-ATP

used in a wide variety of related problems such as linear inverse problems, data compression and statistical estimation (see [3, 4, 5] for references).

In this paper we first give a short description of the classical wavelet denoising (CWD) algorithm, its properties and shortcomings. We then describe a new algorithm, the undecimated (shift invariant) wavelet denoising (UWD) algorithm, and give its properties. Most good properties of CWD carry over and several problems are resolved. For various test signals and signal to noise ratios (SNR) UWD performs considerably better than CWD, both visually and in the l_2 sense. We were first introduced to the idea of redundant denoising in conversations with R. Coifman, and we have more recently seen the ideas discussed here in other publications [5, 9]. However, to our knowledge the combination of the shift invariant discrete wavelet transform [10] and nonlinear processing [5] in order to perform denoising is analyzed here for the first time.

Denoising by thresholding — a review

This section summarizes the results of Donoho and Johnstone [3, 4, 5]. Let $y_i = x_i + \sigma n_i$, $i = 1, \ldots, N$ be a finite length signal of observations of the signal x_i that is corrupted by i.i.d. zero mean, white Gaussian noise n_i with standard deviation σ, i.e., $n_i \overset{iid}{\sim} \mathcal{N}(0, 1)$. The goal is to recover the signal x from the noisy observations y. Here and in the following, v denotes a vector with the ordered elements v_i if the index i is omitted. Let W and M denote the discrete wavelet transform (DWT) matrix and its inverse, respectively. Capital letters symbolize the DWT of a signal, i.e., $X = Wx$, $x = MX$. Furthermore, let $\widehat{X}$ denote an estimate of X, based on the observations Y. The diagonal linear projections $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_N)$, $\delta_i \in \{0, 1\}$, $i = 1, \ldots, N$ give rise to the estimate

$$\hat{x} = M\widehat{X} = M\Delta Y = M\Delta W y, \tag{1}$$

i.e., keeping or killing the individual wavelet coefficients. The risk measure is defined as

$$\mathcal{R}(\widehat{X}, X) = E\left[\|\hat{x} - x\|_2^2\right] = E\left[\|M(\widehat{X} - X)\|_2^2\right] = E\left[\|\widehat{X} - X\|_2^2\right], \tag{2}$$

where the last equality is valid only in case of an orthogonal W (e.g., $M = W^T$). The optimal coefficients in the diagonal projection scheme are $\delta_i = 1_{|X_i| > \sigma}$, i.e., only those values of Y where the corresponding elements of X are larger than σ are kept, all others are set to zero. The corresponding risk $\mathcal{R}_{id}(\widehat{X}, X) = \sum_{n=1}^{N} \min(\widehat{X}_n^2, \sigma^2)$ is called the ideal risk.

Donoho and Johnstone mainly consider two thresholding schemes, hard thresholding (keep Y_i if it is above some threshold τ, set it to zero else) and soft thresholding (additionally shrink those values Y_i by τ that are not set to zero). Among several interesting properties, the following are especially important: Both schemes are within a logarithmic factor ($\log N$) of the ideal risk. Hard thresholding typically yields a smaller mean square error (MSE). Soft thresholding achieves near minimax MSE subject to the constraint that $\hat{x}$ is (with high probability) at least as smooth as x. In this scheme N values have to be stored and the computational complexity is $O(N)$.

The practical issues are: Hard thresholding exhibits spurious oscillations; soft thresholding avoids spurious oscillations. Similar to classical denoising methods (e.g., low pass filtering) there is a tradeoff between noise reduction and oversmoothing of signal details. Since the DWT is not shift invariant, CWD is not shift invariant. Hence, the denoising performance can significantly change by changing the initial shift of the noisy signal.

2 The method

A more detailed description of the proposed method can be found in [7]. Shift invariance can be achieved by computing the wavelet transform of *all* shifts and is called shift invariant DWT (SIDWT). It requires the storage of $N \log N$ values and has computational complexity $O(N \log N)$ [1, 8].

The new method is based on thresholding the SIDWT and can be interpreted as averaging the result of CWD for all possible shifts of the input signal. The basic difference is that the corresponding transformation matrix W is not square and consequently not orthogonal. However, a left inverse, M, exists and can be computed with complexity $O(N \log N)$. In fact, the representation in the SIDWT domain contains redundancy that adds to the improved performance (compare discussion on frames in [2]). In contrast to the case where W is orthogonal (for which the noise terms, N_i, are uncorrelated) the nonorthogonal transform yields correlated noise terms. Since the analysis for the CWD assumes orthogonality the above results do not automatically carry over to UWD. We will present some of the important theoretical properties of the UWD in the following.

Performance analysis

The analysis of the ideal risk for the SIDWT is similar to that by Guo [6]. Define the sets $A = \{i \mid |X_i| \geq \sigma\}$ and $B = \{i \mid |X_i| < \sigma\}$. A vector or matrix indexed by A (or B) indicates that only those rows are kept that have indices out of A (or B). All others are set to zero. Using this notation the ideal risk

$$\begin{aligned} \mathcal{R}_{id}(\tilde{X}, X) &= E\left[\|M(\tilde{X} - X)\|_2^2\right] \\ &= \sigma^2 \mathrm{tr}\left[M W_A W_A^T M^T\right] + X_B^T M^T M X_B \end{aligned} \tag{3}$$

can be derived, where $\mathrm{tr}[\cdot]$ denotes the trace. Notice that for orthogonal W (3) immediately specializes to $\mathcal{R}_{id}(\widehat{X}, X) = \sum_{n=1}^{N} \min(\widehat{X}_n^2, \sigma^2)$, given above.

Properties

Smoothness: Smoothness of the estimate $\hat{x}$ for UWD can be guaranteed in the same way as for CWD. The argument is as follows: Apply soft thresholding to the SIDWT (this guarantees smoothness for all possible shifts of the estimate). It is also easy to show that averaging preserves smoothness since smoothness spaces are vector spaces.
Observations: Hard tresholding with slightly increased threshold yields smooth estimates with low l_2 error. This is in contrast to CWD where one has to sacrifice one for the other of these properties.

Design parameters

In this paper we do not deal with the possible improvements resulting from optimally choosing the actual parameters. We rather consider the same wavelet filter (Daubechies), the same number of levels and an optimal threshold for both the DWT and the SIDWT. Additional improvements are possible [7] by optimizing over all free parameters which is, in fact, the major concern of most of the recent publications on classical wavelet denoising.

3 Example

In our example we use the following set of parameters. The scaling filter is a Daubechies filter of length 6, the signals are of length 512, and the number of scales (or filter bank stages) used is 7. The threshold is chosen as the product of the median absolute deviation (used to estimate the standard deviation [4]) and a fixed number, i.e., 3 for the Donoho method (soft thresholding) and 3.6 for the new method (hard thresholding). The test signal is generated by Donoho's MATLAB routine `MakeSignal` from his software package `WaveLab`.

Fig. 1 shows the result of denoising the signal 'Doppler' plus noise for a signal to noise ratio of 6dB. The result of UWD is clearly better, visually, in terms of smoothness, and in the l_2 sense.

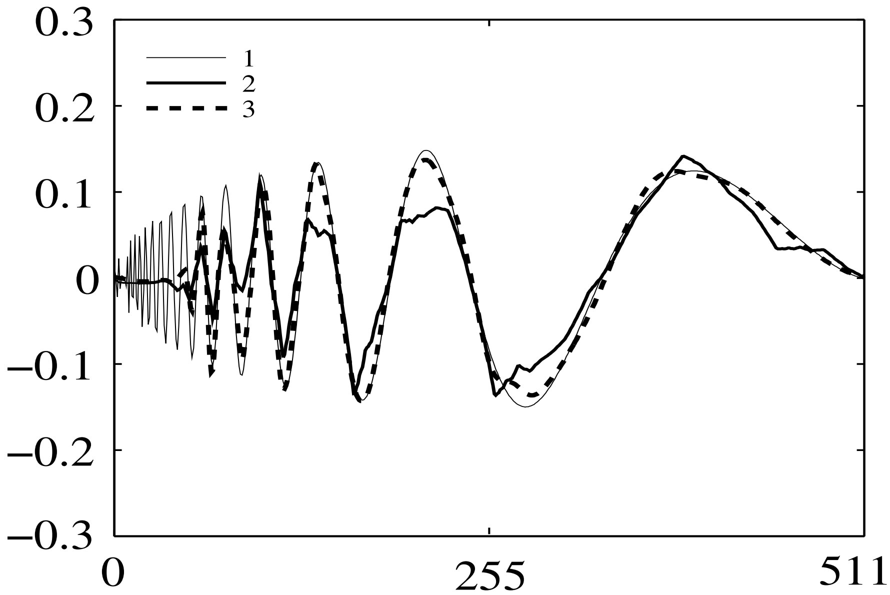

Figure 1: Denoising of signal 'Doppler' for an SNR of 6dB; noise free signal (line 1), result of classical wavelet denoising (line 2), result of undecimated wavelet denoising (line 3).

Fig. 2 depicts the resulting SNR after denoising (SNR_{out}) for different SNR of the noisy signal (SNR_{in}). It is apparent that the SNR corresponding to the ideal risk of the UWD is better than that of CWD by 2–3dB. Furthermore, the SNR of the actual risk follows that of the ideal risk closely and is uniformly better than that of CWD. The difference even increases with increasing SNR. The latter fact is a consequence of soft thresholding. If we had used hard thresholding for the CWD the difference would have been relatively constant (still with UWD performing better than CWD) but the denoised function would have exhibited many small spikes (spurious oscillations).

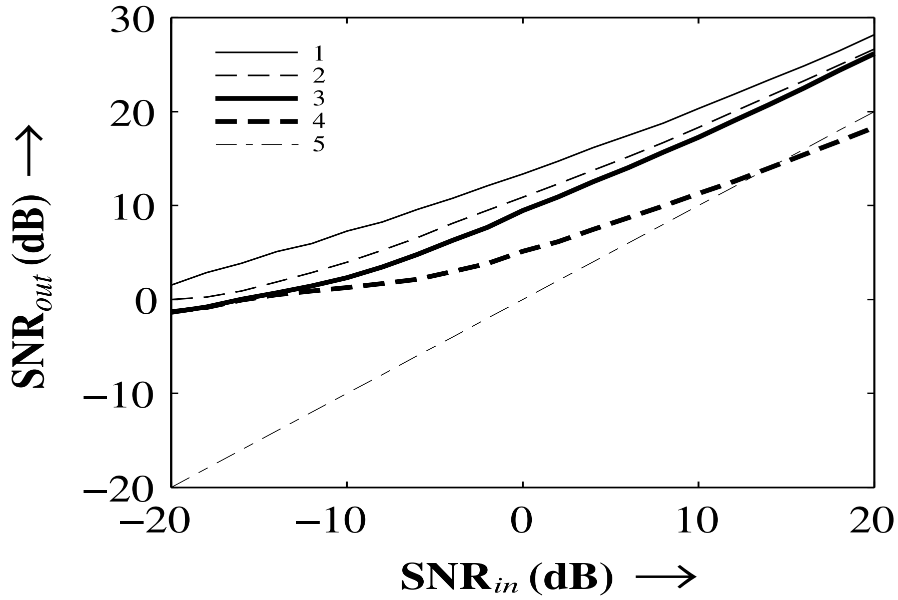

Figure 2: Improvement of the actual and ideal risks by using undecimated wavelet denoising for test signal 'Doppler'. SNR of noisy signal (line 5), resulting SNR$_{out}$ of classical wavelet denoising (line 4), resulting SNR$_{out}$ of undecimated wavelet denoising (line 3), ideal SNR$_{out}$ corresponding to the ideal risk of classical wavelet denoising (line 2), ideal SNR$_{out}$ corresponding to the ideal risk of undecimated wavelet denoising (line 1).

4 Conclusions

We have presented a new denoising algorithm that employs hard thresholding of the shift invariant DWT. It can be interpreted as a generalization of Donoho and Johnstone's method which employs an orthogonal transform. Similar results can be derived such as the smoothness property or a closed form expression for the ideal risk. In contrast to CWD, UWD is shift invariant. Also, in contrast to CWD, smooth and accurate estimates can be computed simultaneously. This is confirmed by an example and holds according to our experience for several applications mentioned in the introduction as well as for the all test functions used in [4]. The required storage as well as the computational load, however, is increased by a factor of $\log N$

compared to the CWD. A MATLAB tool-box for wavelet design and denoising can be obtained from http//www-dsp.rice.edu/

Acknowledgments

We thank the reviewers for comments and suggesting additional references. We also thank R. Coifman for suggesting the method of redundant denoising. Finally we thank all those authors who made their technical reports and publications readily available on the Internet and the World Wide Web.

References

[1] G. Beylkin. On the representation of operators in bases of compactly supported wavelets. *SIAM J. Numer. Anal.*, 29(6):1716–1740, 1992.

[2] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992. Notes from the 1990 CBMS-NSF Conference on Wavelets and Applications at Lowell, MA.

[3] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 41(3):613–627, May 1995.

[4] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994. Also Tech. Report 400, Department of Statistics, Stanford University, July, 1992.

[5] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *J. R. Statist. Soc. B.*, 57(2):301–337, 1995.

[6] H. Guo. Theory and applications of the shift-invariant, time-varying and undecimated wavelet transform. Master's thesis, Rice University, Houston, TX, May 1995.

[7] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr. Nonlinear processing of a shift-invariant DWT for noise reduction. In *SPIE conference on wavelet applications*, volume 2491, Orlando, FL, April 1995. Also Tech. report CML TR95-03, Rice University, Houston, TX.

[8] S. Mallat. Zero-crossings of a wavelet transform. *IEEE Trans. Inform. Theory*, 37(4), July 1991.

[9] G. P. Nason and B. W. Silverman. Stationary wavelet transform and some statistical applications. Technical report, Department of Mathematics, University of Bristol, Bristol, U.K., February 1995.

[10] J. C. Pesquet, H. Krim, and H.Carfantan. Time invariant orthonormal wavelet representations. *To appear in IEEE Trans. SP.*